September, 1993

LIDS-P 2199

Research Supported By:

Draper Laboratory agreement DL-H-467133
AFOSR grant 92-J-0002
ARO grant DAAL03-92-G-0115
Draper Laboratory Research Fellowship

A Wavelet Packet Approach to Transient Signal Classification*

Learned, R.E.
Willsky, A.S.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 1993		**00-09-1993 to 00-09-1993**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Wavelet Packet Approach to Transient Signal Classification	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Laboratory for Information and Decision Systems,77 Massachusetts Avenue,Cambridge,MA,02139**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**35**	

LIDS-P-2199

A Wavelet Packet Approach to Transient Signal Classification[1]

Rachel E. Learned[2]

Alan S. Willsky

September 27, 1993

This paper has been submitted to *Applied and Computational Harmonic Analysis*

Abstract

Time-frequency transforms, including wavelet and wavelet packet transforms, are generally acknowledged to be useful for studying non-stationary phenomena and, in particular, have been shown or claimed to be of value in the detection and characterization of transient signals. In many applications time-frequency transforms are simply employed as a visual aid to be used for signal display. Although there have been several studies reported in the literature, there is still considerable work to be done investigating the utility of wavelet and wavelet packet time-frequency transforms for *automatic* transient signal classification. In this paper we contribute to this ongoing investigation by exploring the feasibility of applying the wavelet packet transform to automatic detection and classification of a specific set of transient signals in background noise. In particular, a noncoherent wavelet-packet-based algorithm specific to the detection and classification of underwater acoustic signals generated by snapping shrimp and sperm whale clicks is proposed. We develop a systematic feature extraction process which exploits signal class differences in the wavelet packet transform coefficients. The wavelet-packet-based features obtained by our method for the biologically generated underwater acoustic signals yield excellent classification results when used as input for a neural network and a nearest neighbor rule.

[1]The work of the first author was supported by the Charles Stark Draper Laboratory under a research fellowship. The work of the second author was supported in part by the Draper Laboratory IR& D Program under agreement DL-H-467133, by the Air Force Office of Scientific Research under Grant AFSOR-92-J-0002, and by the Army Research Office under Grant DAAL03-92-G-0115.

[2]Massachusetts Institute of Technology, Room 35-439, 77 Massachusetts Ave, Cambridge, MA 02139. e-mail: learned@mit.edu, tel: 617-253-6172, fax: 617-258-8553

1 Introduction

Signals possessing non-stationary information are not suited for detection and classification by traditional Fourier methods. An alternate means of analysis needs to be employed so that valuable time-frequency information is not lost. The wavelet packet transform is one such time-frequency analysis tool. This paper examines the feasibility of using the wavelet packet transform in automatic transient signal classification through the development of a simple noncoherent feature extraction procedure for biologically generated underwater acoustic transient signals in ocean noise.

The ability to classify underwater acoustic signals is of great importance to the Navy. Today, detection and classification, tailored for stationary signals, is done by Naval personnel who listen to incoming signals while viewing computer generated displays such as time vs. angle-of-arrival and time vs. frequency. The signal of interest is monitored and the primary frequencies contained in the signal are noted. An initial guess as to the source is made. In efforts to confirm or contradict the guess, the Naval officer will, perhaps repeatedly, consult tables which contain the frequency information on a large range of possible signals.

Transient signals, lasting only a fraction of a second, are of particular concern because they will typically appear as broadband energy on the frequency display. Thus, the Naval officer cannot rely on any visual displays for assistance in the classification process. At present the human observer must be able to detect and classify transient signals by only listening for them. These brief signals may be missed by the listener. An automatic method of classification for transient signals would greatly aid in the detection/classification process.

The frequency display which uses standard spectral analysis methods is useful for stationary signal classification; transient signals are not well matched to these methods. In particular, Fourier-based methods are ideally suited to the extraction of narrow band signals whose durations exceed or are at least on the order of the Fourier analysis window length. That is, Fourier analysis, particularly the short-term Fourier transform (STFT), does an excellent job of *focusing* the information for sources of this type, thus, providing features (spectral amplitudes) perfectly suited to detection and discrimination. The STFT does allow for some temporal as well as frequency resolution, but it is not well suited for the analysis of many transient signals and, in particular, to the generation of features for detection and discrimination.

The STFT may be viewed as a uniform division of the time-frequency space. It is calculated for consecutive segments of time using a predetermined window length. The accuracy of the STFT for extracting localized time/frequency information is limited by the length of this window relative to the duration of the signal. If the window is long in comparison with the signal duration there will be time averaging of the spectral information in that window. On the other hand, the window must be long enough so that there is not excessive frequency distortion of the signal spectrum. The STFT with its non-varying window is not readily adaptable for capturing signal-specific characteristics.

In contrast, the wavelet packet transform offers a great deal more freedom in dealing with this time-frequency trade-off. Indeed, the development of wavelet transforms [2, 7, 8, 10, 11]

and wavelet packets [1, 15] has sparked considerable activity in signal representation and in transient and non-stationary signal analysis. In this paper we are particularly interested in the research that has dealt with automatic detection and classification of transients. These works can roughly be grouped into two categories. One group of methods has focused on problems in which the classes of transients to be detected are well characterized by prior parametric models that identify the distinguishing characteristics of each class. Such methods generally operate based on coherent processing, i.e. on using wavelets as the basis for detection procedures that resemble matched filtering. In particular, Friedlander and Porat [5] find the optimal detector via the generalized likelihood ratio test for three linear time-frequency transforms of the received signal which is characterized by a signal model and a mismatch error in additive white Gaussian noise. They examine the performance of their detector with the short-term Fourier transform, the Gabor transform, and the wavelet transform. Frisch and Messer [6] also formulate a detector by using the generalized likelihood ratio test for the wavelet transform coefficients of the received signal model. They restrict their signal model to an unknown transient with known relative bandwidth and time-bandwidth product. This assumption greatly reduces the complexity of the detector.

The second set of techniques, into which this research falls, deals with the detection and classification of transient signal classes that are not well-characterized in terms of prior models; consequently, somewhat different methods of detection and classification must be developed. In particular, recent work in the area of underwater acoustic transient classification using wavelet related concepts has been done by Lemer, Nicolas, and Legitimus [13] and, more recently, Desai and Shazeer [3]. Both [13] and [3] employ a wavelet packet transform as a means of generating features from various classes of underwater acoustic transients for input to a neural network. The authors of [13] use the energy in the wavelet decomposition of the transients along with features derived from autoregressive signal models and histograms of the data. The authors of [3] use the eight signals resulting from the third level of the wavelet packet decomposition, i.e. each transient signal is separated into eight components, one corresponding to each of eight equal bandwidth channels. The Fourier transform and curve length of these eight sequences are used as features.

One characteristic common to both of these efforts is that the choice of the wavelet packet basis is not considered as part of the feature selection process, consequently, exploitation of class dependent frequency characteristics are suppressed by using a predetermined wavelet packet basis. A natural direction extending beyond these efforts is to address the issue of finding a wavelet-packet-based feature set that offers maximum feature separability due to class-specific characteristics. Our work explores the utility of the wavelet packet transform as a tool in the search for features that may be used in the detection and classification of transient signals in background noise. In particular, we formulate a systematic method of determining wavelet-packet-based features that exploit class-specific differences of the signals of interest.

This paper is organized as follows. Section 2 summarizes wavelet packet notation and establishes the energy mapping of the wavelet packet transform used in this paper. The Charles Stark Draper Laboratory and the Naval Undersea Warfare Center furnished an

extensive collection of acoustic signals in background noise which allowed for an empirical study of some typical occurrences of snapping shrimp and whale clicks. These data are discussed in Section 3. Section 4 details our systematic method for determining wavelet-packet-based features by the formulation of a wavelet-packet-based feature set for snapping shrimp and whale clicks. The focus of our method is on the enhancement of class-specific differences obtained through careful examination of the feature separation attainable from the wavelet packet decomposition of the transients. Using the features from Section 4 with a nearest neighbor rule and a neural network we obtain 98% to 99% classification. These test and results are summarized in Section 5. Other data are analyzed and tested in Section 5.2. Section 6 offers concluding remarks and a discussion of possible future work.

2 The Wavelet Packet Transform and its Energy Mapping

In this section we briefly review the structure of the wavelet packet decomposition (WPD) that was developed by Coifman and Wickerhauser in [1]. We also introduce the notation and quantities to be used in the rest of this paper. The WPD, which can be viewed as a natural extension of the wavelet transform, provides a level by level transformation of the signal from the time domain to the frequency domain. The top level of the WPD is the time representation of the signal. As each level of the decomposition is calculated there is a decrease in temporal resolution and a corresponding increase in frequency resolution.

Using the notation in [1], let $h(n)$ and $g(n)$ be the finite impulse response lowpass and highpass filters used for the decomposition, where the Daubechies 14-point filters [2] are used for all the wavelet packet decompositions in this work. Let $x(n)$ denote the original signal which is of finite length N, where N is a power of 2.

Let F_0 and F_1 be the operators which perform the convolution of $x(n)$ with $h(n)$ and $g(n)$, respectively, followed by a decimation by two. For example, let $x_s(n)$ and $x_d(n)$ denote the sequences resulting from the lowpass filter-decimation operation and highpass filter-decimation, respectively. We have

$$x_s(n) = F_0\{x(k)\} = \sum_k x(k)h(2n-k)$$

$$x_d(n) = F_1\{x(k)\} = \sum_k x(k)g(2n-k).$$

Due to the decimation, $x_s(n)$ and $x_d(n)$ each contain half as many samples as $x(n)$. As Coifman and Wickerhauser do in [1], we also use the s and d notation here because the lowpass F_0 operation may be compared to a sum and the highpass F_1 operation may be compared to a difference.

The wavelet decomposition may be calculated using a recursion of these filter-decimation operations. Figure 1 shows a WPD tree for a signal of length eight. The full WPD is displayed as a tree with a discrete sequence at every branch. Each branch sequence is referred to as a

bin vector for the remainder of this paper. The decomposition may be continued down to the final level where there is only one element in each bin vector. Note that each bin vector is the result of a linear operation (successive applications of the convolution-decimation operators) on the original sequence.

An intuitively pleasing way to view the wavelet packet decomposition tree is by displaying the bins at a given level so that they occur in increasing frequency order from left to right. The method of decomposition described above does not result in a WPD tree displayed in this intuitively pleasing manner. Aliasing occurs which exchanges the frequency ordering of some branches of the tree. A simple swapping of the appropriate bins corrects the problem. All decomposition trees as well as Figure 1 have been rearranged to reflect the "intuitive" frequency ordering of bins.

The bin locations within a tree will be represented by the notation b(l,c) where a bin is indexed by two parameters, level, l, and column, c. Figure 2 shows each bin of a WPD tree labeled with the appropriate bin position notation. For example, the notation b(1,1) refers to the bin at the top level containing the time domain signal. This bin is at level 1 and, since there is only one column at the top level, column 1.

A few examples will illustrate the display we use for the WPD of a signal and the time-frequency trade-off inherent in the WPD. Each bin vector of the WPD tree is displayed as a rectangular intensity plot at its appropriate position in the tree. The magnitude of each element of a bin vector is displayed with black corresponding to the maximum absolute value in the tree and white corresponding to zero.

We begin with Figure 3, a signal comprised of two sinusoids. From Figure 3 we see that as the levels of the WPD tree are traversed, the information becomes more focused. The lowest level of the tree essentially agrees with the discrete Fourier transform of the signal. Shown in Figure 4 is a time and frequency localized signal corresponding exactly to one of the wavelet packet basis functions. Note the focusing of information at bin(5,6) of the tree. The information is less focused at the top and bottom of the tree, thus, the most compact or focused representation would be at bin b(5,6) of the WPD tree.

Two points about these examples are worth noting. First, recall that the *wavelet* transform corresponds to a very particular set of bins, namely those corresponding to successive lowpass/decimation (F_0) operations followed by a *single* highpass/decimation (F_1) operation. As pointed out in [1], only certain types of signals are well-focused in these bins. For example, the signal in Figure 4 is focused at bin(5,6) which is *not* part of the ordinary wavelet decomposition. Second, the principle idea that we wish to exploit in finding useful features for transient detection and classification is precisely this focusing property, i.e. transients with different time-frequency characteristics will focus differently. To exploit this property for signals as in Figure 4 we must use the full wavelet packet transform and not simply the wavelet transform.

2.1 Energy Mapping of the Wavelet Packet Decomposition Tree.

In detection terms the formation of the wavelet packet transform can be viewed as a *coherent* processing step, i.e. each sample of each signal in each bin in the full WPD can be viewed as the output of a matched filter tuned to a particular basis function. At the top of the WPD tree these basis functions are simply unit impulses at each successive time instant, and as we move down the WPD tree the basis functions become more resolved in frequency and more highly decentralized in time. The matter to be determined, then, is how we use this tree of coherently processed signals to perform detection. If the signals that we wish to detect are also described coherently, i.e. as weighted linear combinations of WPD basis functions, then we can use a fully coherent system in which we simply take as our test statistics the same weighted linear combinations of the WPD of the received signal. However, in the problems of interest here, we do *not* have such a prior model for the signals to be classified, and, indeed, a fundamental premise is that the variability in these signal classes precludes such a precise representation. A second premise, however, is that the *energy* in the WPD for these signal classes does focus in a robust and useful way. This suggests a second *noncoherent* (i.e. energy-based) processing step after the WPD has been performed. Specifically, in this work we have done a simple energy mapping of the wavelet packet transforms of our data in order to begin the feature extraction process with a rudimentary exploration of signal specific characteristics.

Let $e_{\mathbf{y}}$ denote the energy of a vector $\mathbf{y}$ having N elements. The average energy in $\mathbf{y}$ is

$$e_{\mathbf{y}} = \frac{1}{N}\,\mathbf{y}^T\mathbf{y} \tag{1}$$

An example of this energy mapping is shown in Figure 5 for the WPD tree from Figure 1 where a single energy value has been calculated for each bin vector. The formation of one energy value over an entire bin obviously loses whatever further time resolution there is within each bin vector. For example, at the top of the WPD we are simply calculating total average energy, a classic test statistic in noncoherent processing. Clearly, we can trade-off between fully coherent and fully noncoherent processing by computing *several* average energies over smaller windows within each bin vector: windows of length one correspond to coherent processing and windows equal to bin length to noncoherent processing. Our intent here, however, is to explore the idea of energy detection in the coherently processed WPD, and thus we restrict attention to the use of a single energy value for each bin. As our results show for the application considered in this paper, this restriction still allows us to achieve excellent performance. In other applications, however, one may wish to use windowed energies in order to determine robust features for signal classification; the procedure outlined in this paper is directly applicable in such cases as well.

3 The Data

This paper uses a collection of ocean recordings made available by the Charles Stark Draper Laboratory and the Naval Undersea Warfare Center (NUWC). The data consists of several

hours of naturally occurring biologically generated underwater sounds in ambient ocean noise. The recordings have been lowpass filtered with a cutoff frequency of 5KHz and, subsequently, sampled at 25KHz. (The Nyquist sampling rate is 10KHz.) The biologically generated sounds are sperm whale clicks and snapping shrimp and are used to illustrate our method of feature extraction for classification. A typical whale click will have a duration of approximately 80 to 120 milliseconds and a single snap of a shrimp will have a duration on the order of 1 millisecond. In addition to the signals, each record contains portions of background noise alone. Figure 6 shows excerpts of whale clicks and snapping shrimp in ambient ocean noise.

A single whale click can be encompassed by a 163.8 millisecond or 4096-sample window which also holds one to an uncountably large number of snaps. Figure 7 shows three 163.8 millisecond excerpts from the NUWC recordings. We use 75 of these excerpts for the feature derivation discussed in Section 4 and 240 additional excerpts to run simulations of the classification algorithms discussed in Section 5.

Using our implementation of Wickerhauser's algorithm presented in [15] with the Daubechies 14 point wavelet [2], the first six levels of the wavelet packet transform of each of the 75 data excerpts were calculated. An energy map was calculated for each of these 75 WPD trees. Each energy map contains 63 bin energies. Figure 8 shows the energy maps of the wavelet packet transforms of three data excerpts.

This mapping of the WPD trees shows promising clarification of information. At a glance, one can see a definite difference in the intensity distributions between the three energy maps shown in the figure. A quantitative analysis of the patterns exhibited by the energy maps is discussed in the next section.

4 Choice of an Optimum Reduced Parameter Feature Set

In the formulation of a decision rule, it is desirable to find a feature set which uniquely represents each class of signals. Typically, the feature set uses a greatly reduced number of parameters in comparison with the number of samples used to represent the signal. In this section, a feature set which best separates characteristics specific to each class is derived from the wavelet packet transforms of the three classes of signals.

4.1 Vector Representation of the Energy Maps

An energy map was found from the WPD of each of the 75 excerpts discussed in Section 3. For ease of manipulation, each energy map is represented as an energy vector by assembling the bin energies of an energy map into a column using lexicographic ordering of the bins. We number the bins from one to 63 and create and energy vector, $\mathbf{e}_{t,k}$, for each of our data excerpts. The element $\mathbf{e}_{t,k}[b]$ is the energy from bin number b of the energy map for the k^{th} signal of class t where t = c (click), n (noise), and s (shrimp).

Next, we create a matrix for each signal class by aligning column vectors of the same class. We denote the energy matrix by E_t

$$E_t = \begin{bmatrix} \mathbf{e}_{t,1} & \mathbf{e}_{t,2} & \cdots & \mathbf{e}_{t,M_t} \end{bmatrix} \tag{2}$$

where M_t represents the number of examples for the given class. Thus, E_t is a $63 \times M_t$ matrix, and in our case $M_t < 63$.

4.2 Examination of the Top Six Levels of the Energy Maps

A first step in the analysis of the transients is to quantitatively identify significant features of all energy maps from a given class. This can be done by looking at the singular value decomposition [14] (SVD) of the matrices, E_t.

$$E_t = U \Sigma V^T \tag{3}$$

The 63-element singular vectors, $\mathbf{u}_k$, make up the columns of the 63×63 orthogonal matrix U. The first M_t columns of U span the column space or range of E_t.

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{63} \end{bmatrix} \tag{4}$$

The $63 \times M_t$ singular value matrix, Σ, displays the rank in the first M_t diagonal elements. The rank (or effective rank) of E_t is equal to the number of non-zero (or non-negligible) singular values.

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{M_t} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{5}$$

The row space and nullspace of E_t are defined in the $M_t \times M_t$ matrix, V^T. The information in V^T is not used in the analysis of the energy maps.

For the energy matrices E_c, E_s, and E_n, we found that each had a single dominant singular value. In particular, define the difference ratio, δ_t, between the largest and second largest singular values for each class to be

$$\delta_t = \frac{\sigma_{t,1} - \sigma_{t,2}}{\sigma_{t,1}}, \quad t = c,\ n,\ s \tag{6}$$

Table 1: Difference ratio between the largest and second largest singular values of the three E_t matrices.

Class t	$\sigma_{t,1} \times 10^6$	$\sigma_{t,2} \times 10^6$	δ_t
whale clicks	2221	285	0.87
snapping shrimp	762	79	0.90
background noise	412	43	0.89

and is displayed in Table 1.

These values suggest that there is a single representative energy vector, corresponding to the first singular vector $\mathbf{u}_{t,1}$, for each class, t, with only a relatively small amount of variation across class members.

All 63 elements of the primary singular vector for each class are displayed in Figure 9. Notice that high valued elements for the noise coincide with both the high valued elements for the snapping shrimp and the whale clicks. The figure also reveals that the high valued elements for whale clicks differ from the high valued elements for snapping shrimp. Before continuing the search for a reduced parameter feature vector from the energy maps of the sample signal data set, the influence of noise may be compensated for.

4.3 Compensating for the Noise

Each bin energy, i.e. each component of the energy vector, $\mathbf{e}_{t,k}$, contains both signal and noise energies. The energy maps of background noise displayed consistent energy distribution patterns. An example is seen in the sample energy map of background noise from Figure 8. This distribution of background noise energy within the energy maps may mask dominant features that may be useful in distinguishing between the shrimp and clicks. We wish to normalized each bin energy by an average noise energy so that features may be chosen without the influence of noise.

Let $\mathbf{r}$ denote the portion of the received signal vector that is due to the signal source alone. Let $\mathbf{w}$ denote the portion of the received signal vector that is due to background noise. The received signal vector, $\mathbf{x}$, may be written as a linear combination of the source

signal and the background noise.

$$\mathbf{x} = \mathbf{r} + \mathbf{w} \tag{7}$$

Let $\mathbf{x}_b$ denote the vector at bin b of the WPD of $\mathbf{x}$. We denote the vector at bin b of the WPD of $\mathbf{r}$ and $\mathbf{w}$ as $\mathbf{r}_b$ and $\mathbf{w}_b$, respectively. Because the wavelet packet decomposition is a linear transform, the bin vector at each bin of the WPD tree can be written as a linear combination of the bin vector due to the source and the bin vector due to the noise, i.e.

$$\mathbf{x}_b = \mathbf{r}_b + \mathbf{w}_b. \tag{8}$$

In agreement with (1), the energy due to the bin vector $\mathbf{x}_b$ will be denoted by $e_{\mathbf{x}_b}$. Likewise, the energy in $\mathbf{r}_b$ is denoted by $e_{\mathbf{r}_b}$ and the energy in $\mathbf{w}_b$ denoted by $e_{\mathbf{w}_b}$. Assuming that the noise is uncorrelated with the signal allows us to write the energy at any bin of the WPD tree as a linear combination of the energy due to the source and the energy due to noise.footnoteHere by "uncorrelated" what we are in essence assuming is that the time-averaged product of the noise and signal components over each bin is zero.

$$e_{\mathbf{x}_b} = e_{\mathbf{r}_b} + e_{\mathbf{w}_b} \tag{9}$$

Normalization of the bin energy, $e_{\mathbf{x}_b}$, by the energy in that bin due to noise alone would give $\hat{e}_{\mathbf{x}_b}$.

$$\hat{e}_{\mathbf{x}_b} = \frac{e_{\mathbf{x}_b}}{e_{\mathbf{w}_b}} = \frac{e_{\mathbf{r}_b}}{e_{\mathbf{w}_b}} + 1 \tag{10}$$

Performing the normalization described in the above paragraphs allows for a source-signal-energy to noise-energy ratio analysis of the patterns exhibited by the energy maps.

Recall from Section 4.2 that we found an energy vector, $\mathbf{e}_{t,k}$, for each of our data excerpts. We now wish to find a normalized energy vector, $\hat{\mathbf{e}}_{t,k}$, for each $\mathbf{e}_{t,k}$. Element by element normalization of $\mathbf{e}_{t,k}$ by the average noise energy elements is done by

$$\hat{\mathbf{e}}_{t,k}[b] = \frac{\mathbf{e}_{t,k}[b]}{\mathbf{e}_{w,ave}[b]}, \tag{11}$$

where the element index is $b = 1,...,63$, the signal number is $k = 1,...,M_t$, and each class is denoted by $t =$ c, s, n. The average noise energy for bin b of the energy maps from our noise excerpts is used for the noise energy, $\mathbf{e}_{w,ave}$. Element by element (or bin by bin) calculation of the average noise energy is done by

$$\mathbf{e}_{w,ave}[b] = \frac{1}{M_n} \sum_k \mathbf{e}_{n,k}[b]. \tag{12}$$

As discussed in Sections 4.1 and 4.2, we may align these normalized energy vectors into three matrices and perform singular value decomposition. The effective rank of each of these matrices was also found to be one so that one singular vector may be used as a representative energy map for each class. Figure 10 shows all 63 elements of the singular vectors found from SVD of the noise normalized energy matrices. We see that the high valued elements of the shrimp singular vector differ from the high valued elements of the whale click singular vector and that there is no longer high valued elements for noise.

4.4 Searching the Noise Normalized Energy Maps for a Reduced Feature Set

In forming a reduced parameter feature set, we look for dominant bin energies that will give us the best separation between whale clicks and snapping shrimp. We begin by finding a collection of bins that contain significant information by examination of the components of the primary singular vectors shown in Figure 10. Let us denote these noise normalized singular vectors by $\hat{\mathbf{u}}_{c,1}$ (whale clicks) and $\hat{\mathbf{u}}_{s,1}$ (snapping shrimp).

To make this specific, we have chosen to consider a bin to be significant if the value of its corresponding element of the primary singular vector lies within 20 percent of the maximum component of that singular vector.footnoteIf the number of bins included in this collection needs to be increased, the threshold may be lowered, step by step, until the desired number of bins is chosen. The significant components of $\hat{\mathbf{u}}_{c,1}$ correspond to elements 9, 18 and 19. The significant components of $\hat{\mathbf{u}}_{s,1}$ correspond to elements 8 and 17. The two classes have no dominant bins in common. The bins corresponding to elements 8, 9, 17, 18 and 19 containing the dominant information are shaded in Figure 11.

Reduction of the feature vector is desirable for the simplification of the decision rule, therefore, including superfluous information should be avoided. A feature set which contains a parent bin energy and all of its descendant bin energies may be redundant because any parent bin vector of the WPD tree can be constructed from its children bin vectors. Therefore, a feature set that does not incorporate both parent and child energy bins of the energy map should be considered. Reducing the number of features used for classification will also minimize the computational complexity of the algorithm because most bins of the WPD tree will not be used and will, therefore, not be calculated.

Looking at Figure 11, we see that the dominant bins b(4,1) and b(4,2) are parents to the rest of the dominant bins b(5,2), b(5,3), and b(5,4). Noting the parent child redundancy, it is reasonable first to see if there is enough feature separation using the energies from only the dominant parent bins at the fourth level, b(4,1) and b(4,2). These two bins are shaded in Figure 12. Figure 13 plots the normalized energies from bins b(4,1) and b(4,2) for the energy maps of our 75 data excerpts. There is excellent separation between the click and shrimp features.

5 Tests and Results

5.1 Original Data Set

Once a reduced parameter feature set has been derived for a given set of data classes, a method for detection and classification must be formulated. In testing the utility of the wavelet-packet-based feature[11 s, we used two pattern recognition techniques that lend themselves to the classification of signals using a training set. The two different classification techniques, the nearest neighbor rule and neural networks, were tested on the biological

transient data using the wavelet-packet-based features discussed in Section 4. These classification methods are discussed in greater detail in Learned's thesis [9].

The nearest neighbor rule, detailed by Duda and Hart in [4], uses as a training set feature vectors that have been correctly classified. A feature vector is calculated for the unknown signal. The unknown feature vector is classified with the same label as its nearest neighboring feature vector from the training set. Euclidean distance is the measure used in determining separation of feature vectors. The training set comprises a number of sample vectors from each class. For example, let there be two classes of interest, class a and class b. We have n_a sample vectors from class a and n_b from class b. We denote the training set of sample feature vectors as $Z = \{\mathbf{z}_{a,1}, \cdots, \mathbf{z}_{a,n_a}, \mathbf{z}_{b,1} \cdots, \mathbf{z}_{b,n_b}\}$ and the unknown feature vector as $\mathbf{z}$. The estimation of the class of our unknown feature vector is denoted by $t(\mathbf{z}, Z)$. We have

$$t(\mathbf{z}, Z) = \arg \min_t (\min_i \mathbf{z}^T \mathbf{z}_{t,i}), \tag{13}$$

where $t = a, b$.

The neural network tests were done using the Neuralware software package [12] for building, training, and analyzing a layered neural network. A back propagation network with a tanh nonlinearity and the Widrow-Hoff-Delta Rule adaptive weighting algorithm was used in all tests. The number of levels and adaptive linear nodes (ALN) used for the neural networks is detailed along with the results for each test.

As mentioned previously, a total of 75 signal segments (consisting of 29 whale clicks, 20 snapping shrimp excerpts, and 26 segments of noise) were used to determine the bin energies which were to be used as features. These features for this set of 75 examples were then used to establish the nearest neighbor rule and to train the several neural networks that were used. Another distinct 240 excerpts from the same overall data set were then used to test classification performance. Each test was run twice; once with the two-parameter feature set determined in Section 4.4 and once with an eleven-parameter feature set comprised of the five bin energies determined in Section 4.4 and six of their adjacent bins.

The nearest neighbor rule algorithm using both the two-parameter and eleven-parameter feature set resulted in correct classification for 97.92% of the test signals. These results are summarized in Table 2. Both nearest neighbor rule tests resulted in identical results, making the same errors. Gaining nothing by adding more features is not surprising because the analysis done in Section 4 determined that the energies from bins b(4,1) and b(4,2) were the dominant features necessary in distinguishing among the three classes.

Three neural networks were constructed for tests using two and eleven features. Excellent results were obtained for all tests. The networks (the number of ALNs in each layer) and their results are summarized in Table 3. The neural networks did an excellent job; classification ranged from 98.33% to 99.17%. Here, we see that only a slight gain in performance results from the addition of the child bins to the two-parameter feature set.

After training a neural network, the weights for the inputs to each ALN may be examined to see which network inputs were found to be most important in the classification process. Large weights consistently appeared at the same five inputs of each ALN in the first layer.

Table 2: Results obtained from the nearest neighbor rule.

Number of Features	2	11
Overall Classification (%)	97.92%	97.92%
Click Classification (%)	97.87%	97.87%
Shrimp Classification (%)	97.26%	97.26%
Noise Classification (%)	98.63%	98.63%

Table 3: Results obtained from the neural network.

Number of Inputs	2	11	11
Number of ALNs in Layer 1	3	7	7
Number of ALNs in Layer 2	0	0	3
Overall Classification (%)	98.33%	98.75%	99.17%
Click Classification (%)	97.87%	98.44%	98.44%
Shrimp Classification (%)	98.63%	98.63%	100%
Noise Classification (%)	98.63%	98.63%	98.63%

Table 4: The difference ratios of singular values for noise normalized energy matrices that include the new shrimp data.

Class t	$\sigma_{t,1}$	$\sigma_{t,2}$	$\sigma_{t,3}$	$\sigma_{t,4}$	$\delta_t^{1,2}$	$\delta_t^{1,3}$	$\delta_t^{1,4}$
snapping shrimp	65.59	25.25	13.19	9.26	0.615	0.799	0.859
whale clicks	124.20	25.16	16.30	10.44	0.797	0.869	0.916

These correspond to the five bins that were determined significant by the analysis done in Section 4.4.

5.2 Results Including a Second Data Set

Some additional recordings of snapping shrimp that were taken at a different time of day and in a different region of ocean than the shrimp used in the previous sections were available to us. Testing these data with both the original two-feature and eleven-feature nearest neighbor rules and neural networks discussed in the previous section (trained without samples of this new shrimp data) resulted in a higher level of incorrect classification. For the new data, the reason for this can be immediately discerned from the cluster distributions shown in Figure 14. The figure shows the two features (energies from bins b(4,1) and b(4,2)) for both the new shrimp data and the original training data set. Notice that the bin energies taken from the new shrimp data records form a cluster to the left of the noise cluster and are distinctly separated from the bin energies for the first shrimp data set. This suggests that there is more variability in the bin energy patterns than that found in the first data set, requiring a richer set of features to capture this behavior. The question is, of course, whether this can be done in a way that still achieves significant feature separability *between* classes.

To explain these issues, 16 excerpts of new shrimp data were appended to the shrimp matrix and the analysis from Section 4.2 was repeated. The four largest singular values for snapping shrimp (old and new data together) and whale clicks are shown in Table 4. We calculate a difference ratio, $\delta_t^{1,i}$, between the largest and i^{th} largest singular values with $i = 2, 3, 4$ as shown in (14).

$$\delta_t^{1,i} = \frac{\sigma_{t,1} - \sigma_{t,i}}{\sigma_{t,1}} \tag{14}$$

Note that the second singular value for the shrimp class is a significant fraction of the largest singular value, thus, confirming that there is indeed more variability in the energy patterns for snapping shrimp excerpts. What we have done is to expand our set of candidate features by examining two singular vectors for shrimp. Figure 15 shows the primary singular vector for whale clicks, the first singular vector for snapping shrimp, and the second singular vector for snapping shrimp scaled by $\frac{\sigma_{s,2}}{\sigma_{s,1}}$. This scaling was done to show the relative intensity of the two shrimp singular vectors.

We begin the search for a feature set by finding significant elements for each of the three singular vectors. We consider an element to be significant if its magnitude is within 25% of the maximum magnitude for that vector. The thirteen significant values found by this procedure are marked with circles in Figure 15 and correspond to bins at levels four, five and six of the energy maps. From the whale click singular vector, bins b(4,2), b(5,3), b(5,4), b(6,6), and b(6,7) are significant. From the first singular vector for snapping shrimp, bins b(4,1), b(5,2), b(6,3) and b(6,4) are significant. From the second singular vector for snapping shrimp, bins b(5,2), b(6,4), b(6,8), b(6,9), b(6,24), and b(6,25) are significant. These bins are shaded in Figure 16–(a).

We wish to have no child-parent redundant bins in our feature set, hence, we note the following: for whale clicks, b(4,2) is an ancestor to b(5,3), b(5,4), b(6,6), and b(6,7); for shrimp singular vector 1, b(4,1) is an ancestor to b(5,2), b(6,3) and b(6,4); for shrimp singular vector 2, b(5,2) is the parent of b(6,4). Using only the ancestor bins, we are left with seven bin energies in our feature set: b(4,1), b(4,2), b(5,2), b(6,8), b(6,9), b(6,24), and b(6,25). These seven bins are shaded in Figure 16–(b).

We used the nearest neighbor rule and neural networks to test the utility of these features for classifying excerpts that include both the old and new data sets. Each pattern recognition rule was run twice, once with the seven-parameter feature set and once with the thirteen-parameter feature set.

The results of the nearest neighbor rule are summarized in Table 5. Errors made by the 13-input nearest neighbor rule are a subset of the errors made by the 7-input nearest neighbor rule. The nearest neighbor tests using both the 13 and 7 features gave excellent results ranging from 86.30% to 95.74% correct classification.

We have not presented results for neural networks for this set of experiments because of serious problems with convergence to local minima. Indeed, one of the benefits of performing the detailed feature analysis we have described is that it leads to a very small set of features that provide excellent inter-class separation. This, in turn, allows us to use a very simple classification rule, namely nearest neighbor, thus avoiding the convergence problems of neural networks.

6 Concluding Remarks

This work has explored the feasibility of applying the wavelet packet transform to detection and classification of unknown transient signals in background noise, i.e. the signals are not

Table 5: Results obtained from the nearest neighbor rule on second data set.

Number of Features	7	13
Overall Classification (%)	91.06%	95.03%
Click Classification (%)	94.68%	95.74%
Shrimp Classification (%)	91.11%	94.81%
Noise Classification (%)	86.30%	94.52%

well characterized by a signal model. In this paper we have detailed our systematic feature extraction process which exploits signal class differences in the wavelet packet transform coefficients. The wavelet-packet-based features obtained by our method for these biologically generated underwater acoustic signals yield 86% to 100% correct classification when used as input for a neural network and a nearest neighbor rule.

The formulation of a wavelet-packet-based feature set explored here combines the coherent processing of the wavelet packet decomposition with noncoherent energy calculations in each bin. From singular value decomposition of matrices made from the energy maps of these data, we found that only a very small number of features were necessary to distinguish among snapping shrimp, whale clicks, and background noise.

We believe that these results are significant not because they provide a definitive algorithm for biological acoustic transients, but rather because they provide convincing evidence that the wavelet packet transform can be used effectively as the basis for robust feature extraction and automatic identification of transient signals that cannot be well-characterized by parametric signal models. Obviously, there is much more work that can be done to develop these ideas. First, as the results in the preceding section make clear, the development of robust classification rules require the availability of data sets that display the full range of variability present in the signal classes to be distinguished (although, as the results in Section 5.2 demonstrate, even a considerable level of variability may still be captured with comparatively small feature sets – a maximum of 13 in this case). Also, as we have pointed out, a simple extension of the incoherent energy feature calculated for each wavelet packet bin is to use a set of windowed energies for each bin, thereby enhancing temporal resolution

and expanding the set of possible features considerably. The results presented here would seem to indicate that such an extension might lead to only marginal performance improvement for the application considered in this paper, but such enhanced temporal resolution may be of considerable value in other applications such spread spectrum communication and active sonar.

7 Acknowledgement

We wish to express our gratitude to Professor George Verghese of the Massachusetts Institute of Technology for his contribution to this project during the early stages of the development of our feature extraction method.

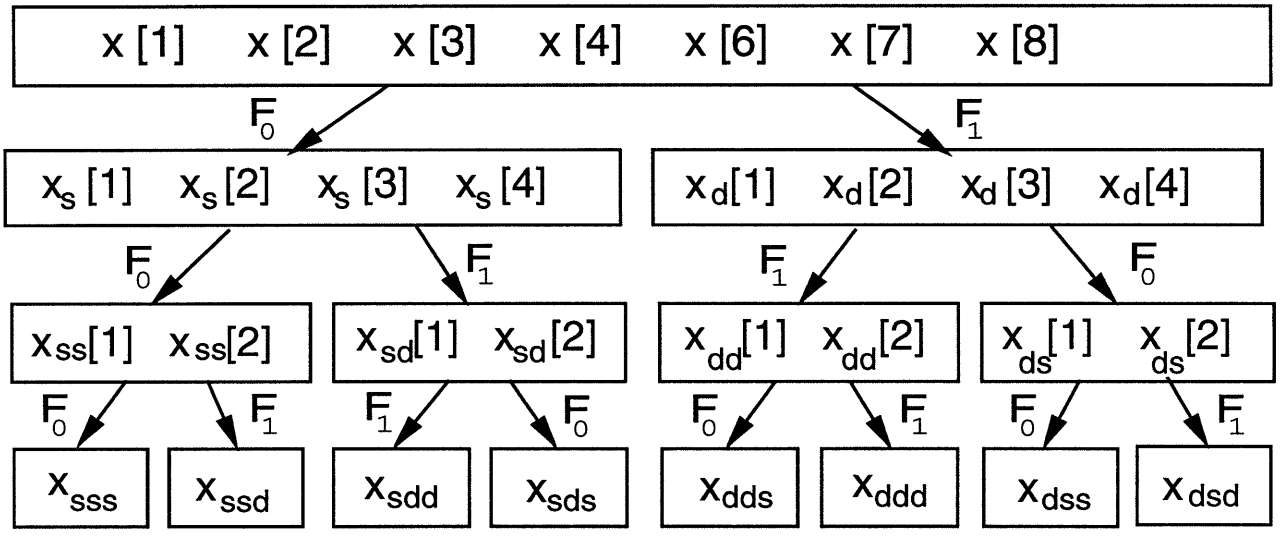


Figure 1: The fully decomposed wavelet packet tree for a signal of length eight.

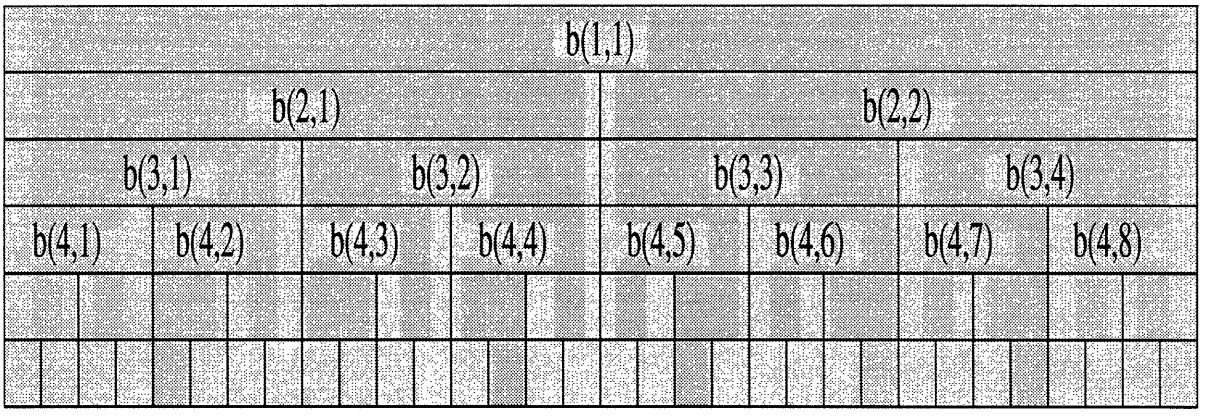


Figure 2: The WPD tree with index label at each bin in the first four levels.

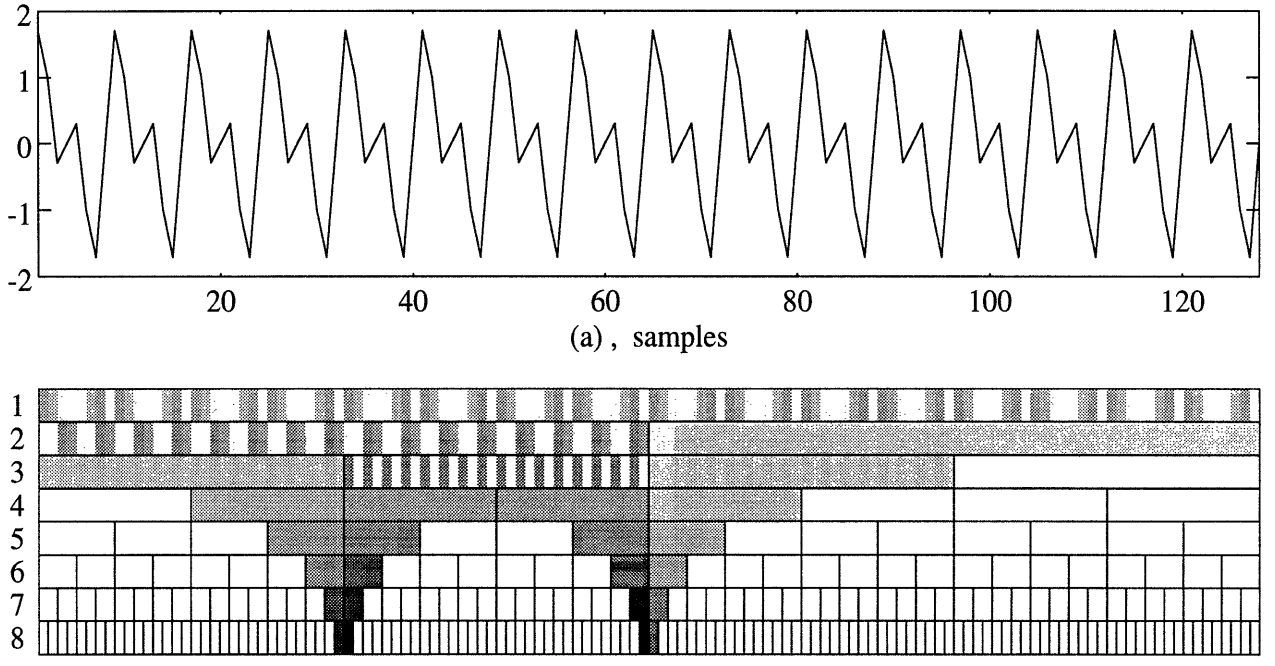


Figure 3: The WPD of two sinusoids. (a) A frequency localized signal (b) WPD of signal.

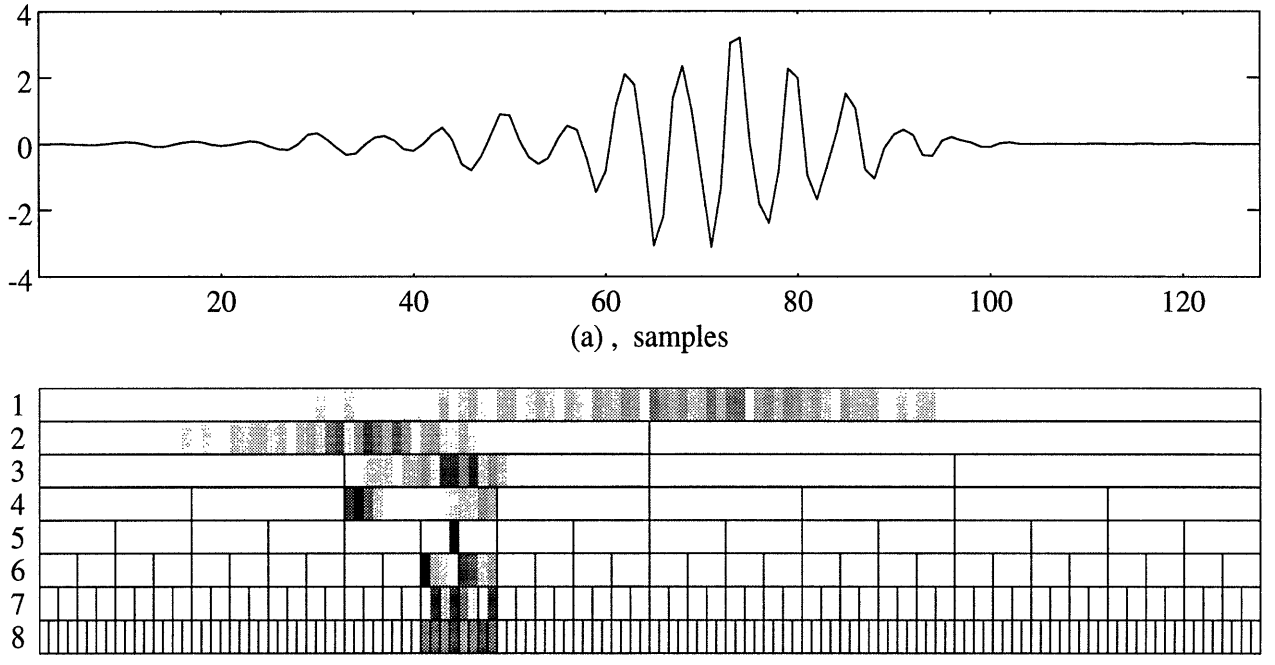


Figure 4: The WPD of a time and frequency localized function. (a) The signal in time (b) WPD of signal

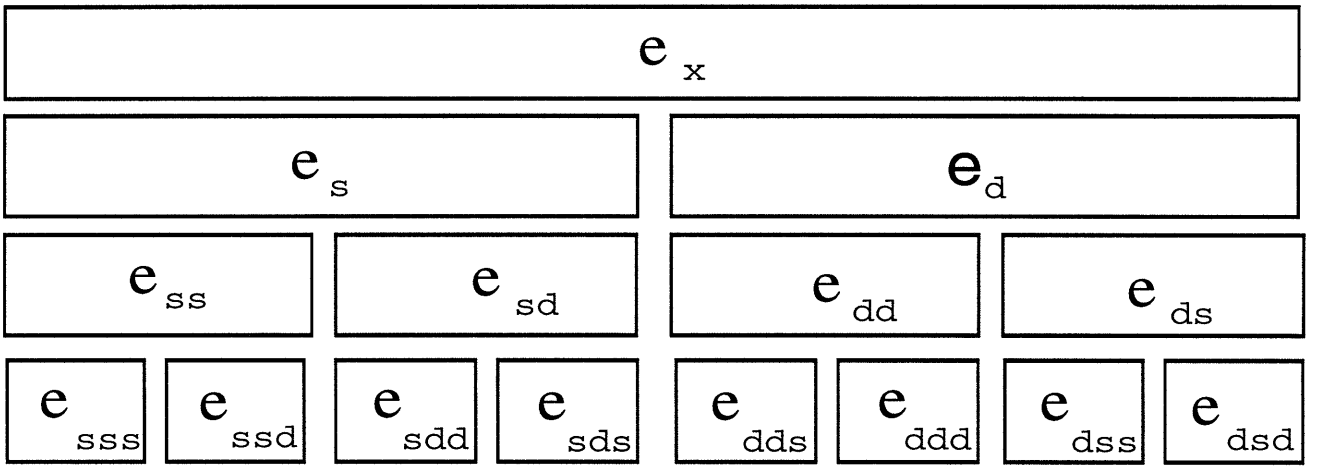


Figure 5: Energy mapping of the WPD tree from Figure 1.

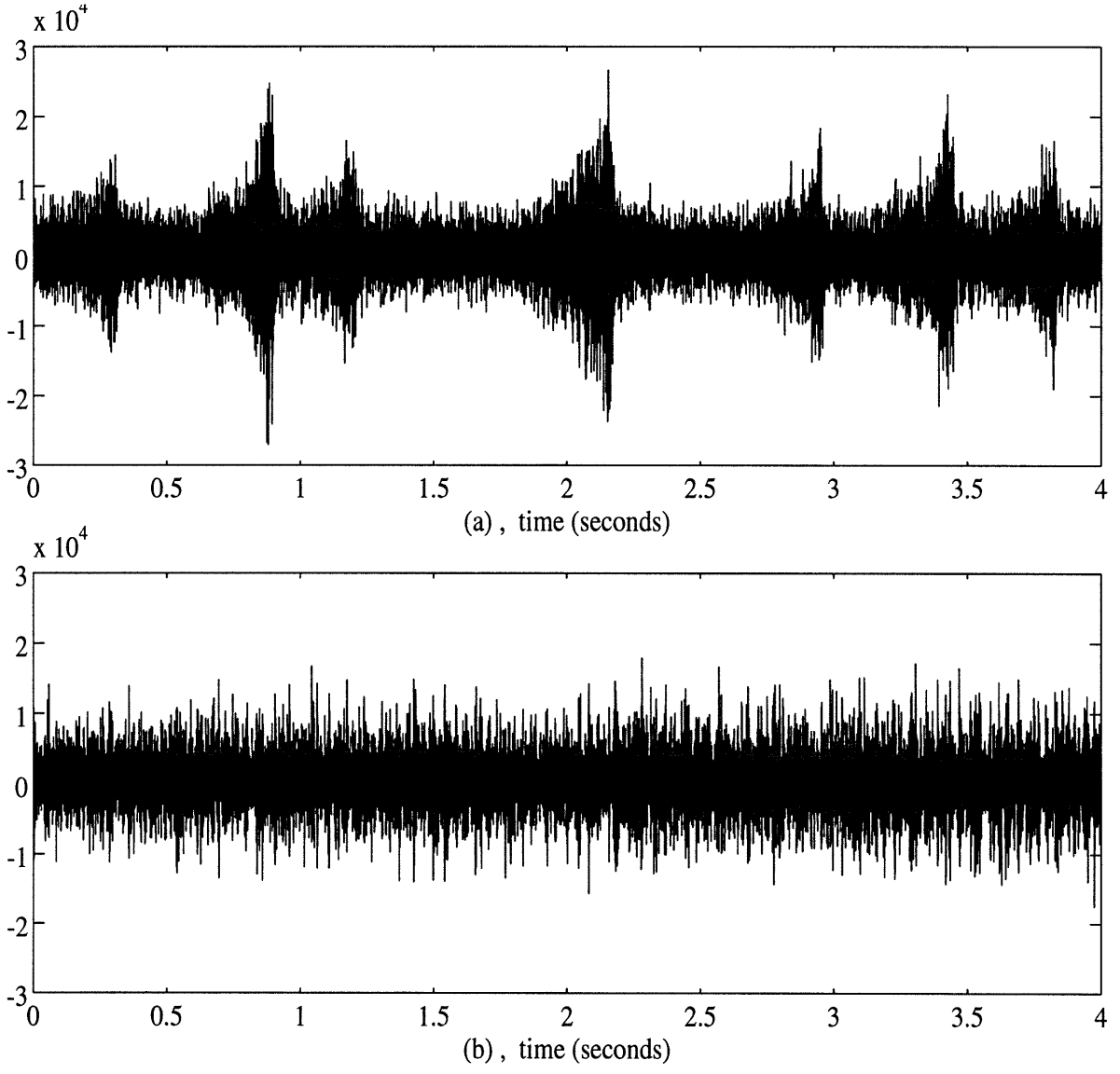


Figure 6: A four second interval of (a) whale clicks and (b) snapping shrimp.

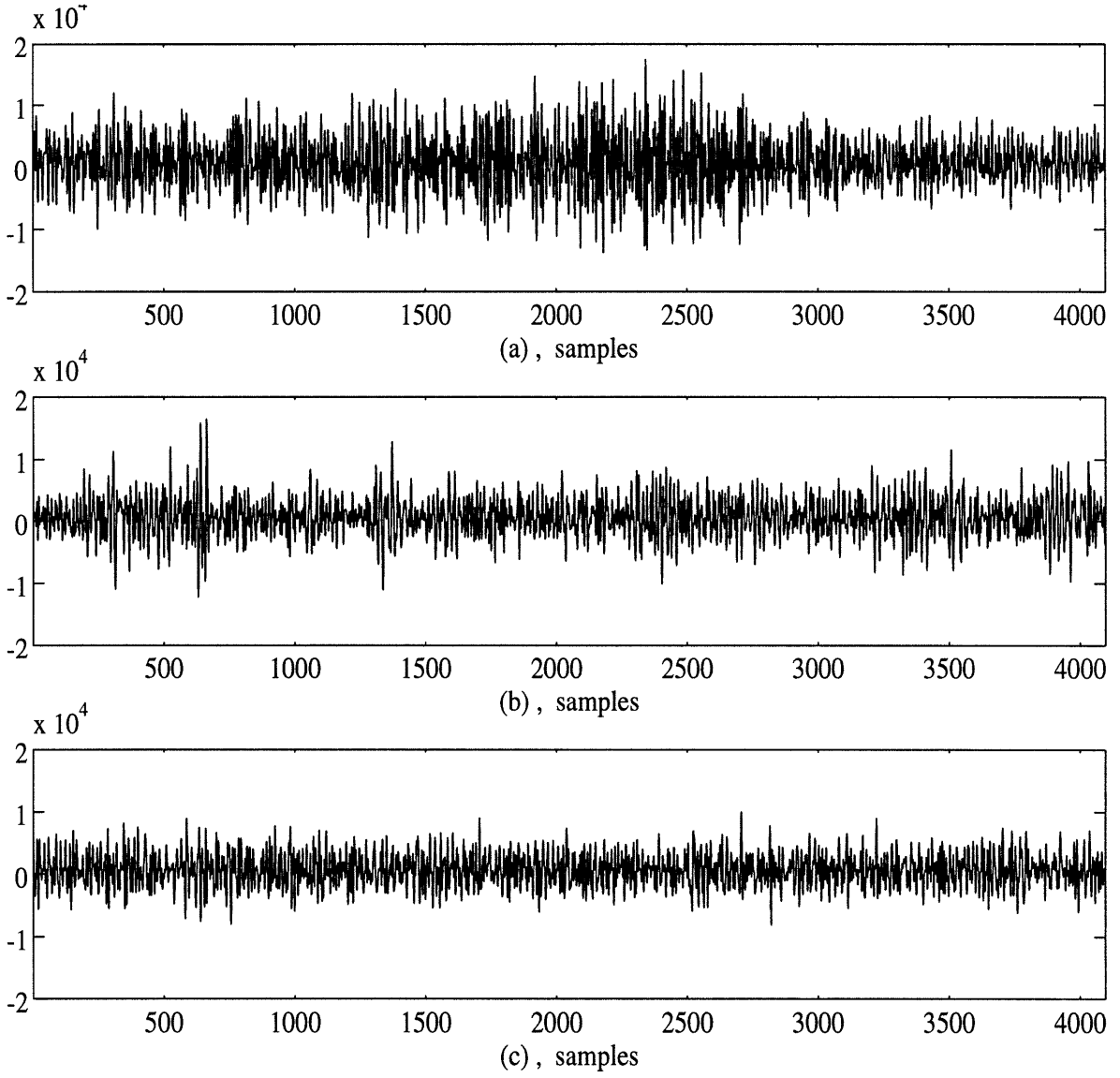
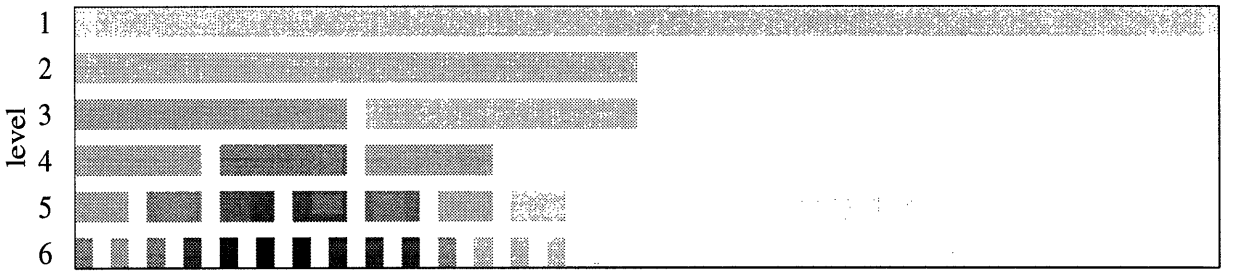
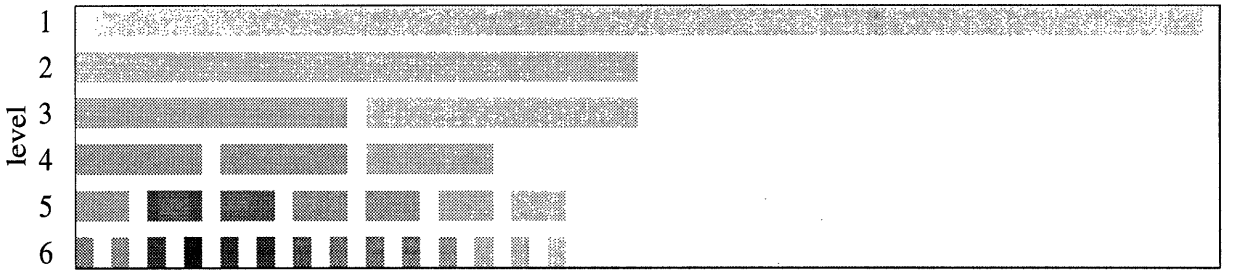


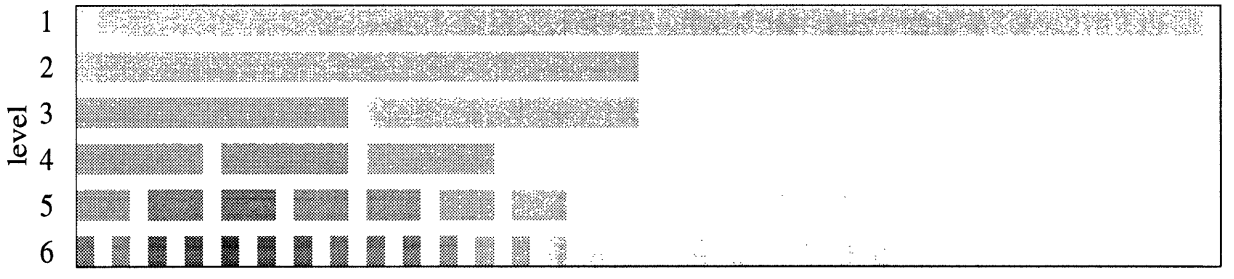
Figure 7: Some 4096-sample (163.8 ms) excerpts from the NUWC recordings. (a) Whale Click. (b) Snapping Shrimp. (c) Background Noise.



(a)



(b)



(c)

Figure 8: Energy maps of the first 6 levels of the wavelet packet transforms of (a) Whale Click (b) Snapping Shrimp (c) Background Noise.

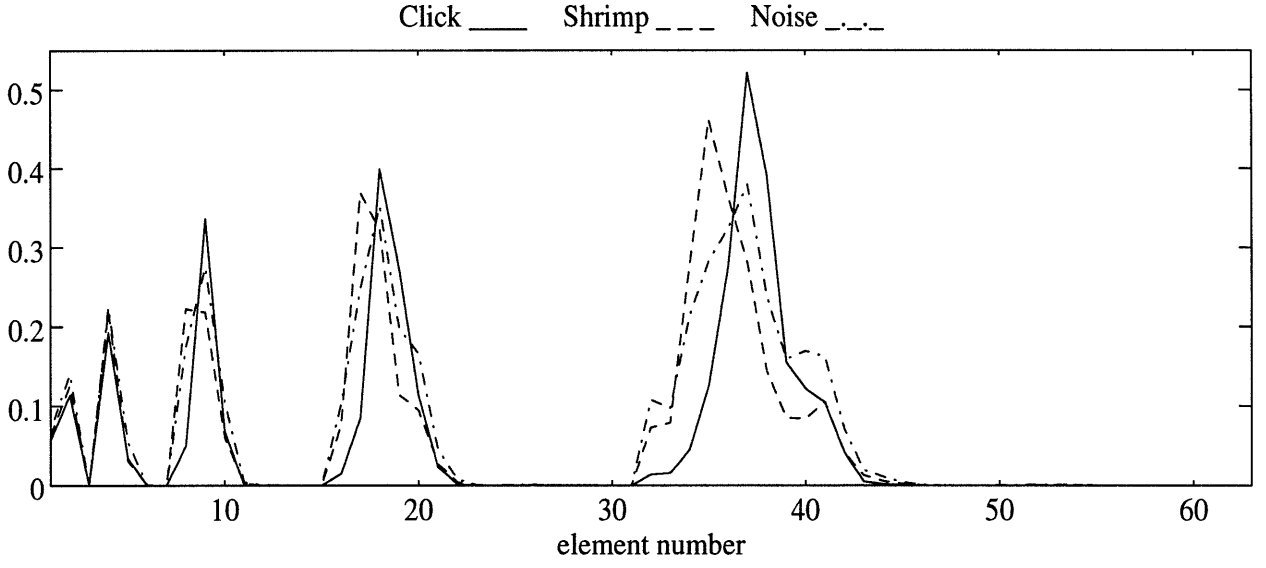


Figure 9: Components of the 63-element primary singular vectors. Note that this plot displays elements of the three primary singular vectors with each vector representing the energy maps for its class. In other words, this plot is a lexicographical display of the representative energy bins for the three classes of energy maps.

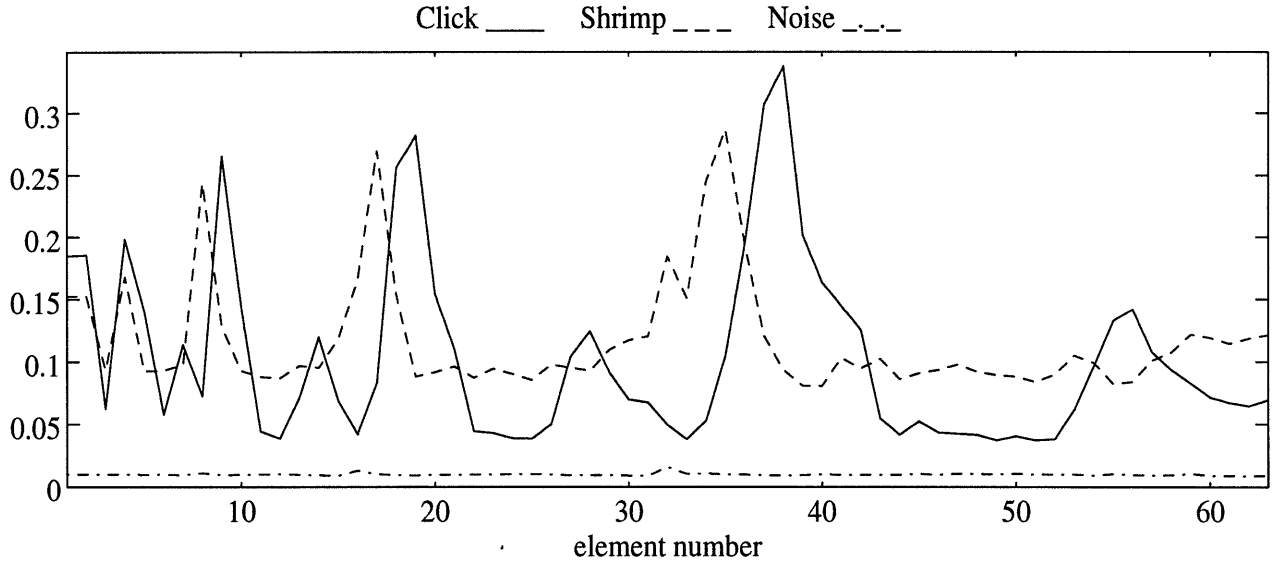


Figure 10: Components of the 63-element primary singular vectors for the noise normalized case.

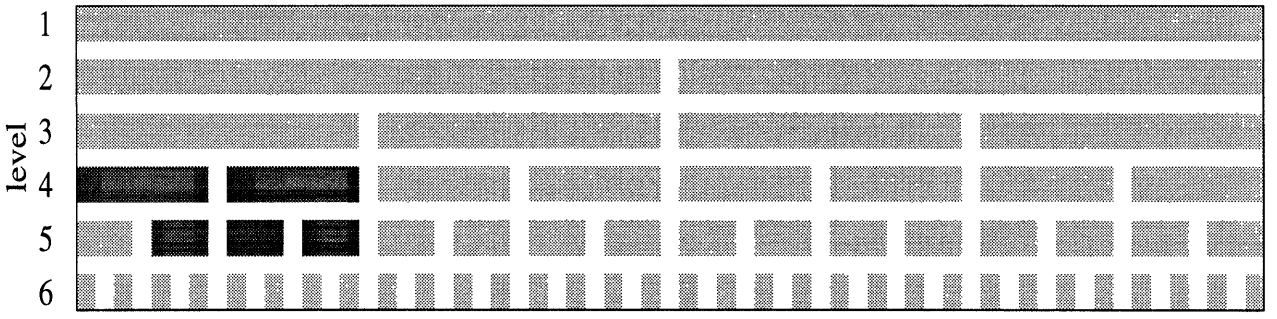


Figure 11: The shaded bins of the energy map correspond to the dominant elements of the primary singular vectors $\hat{\mathbf{u}}_{c,1}$ and $\hat{\mathbf{u}}_{s,1}$.

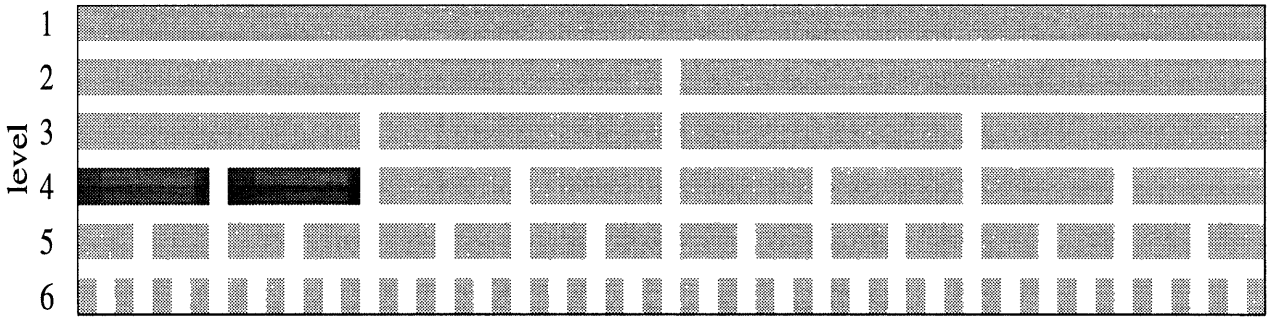


Figure 12: The feature set is composed of the normalized energies from bins b(4,1) and b(4,2) of the energy map.

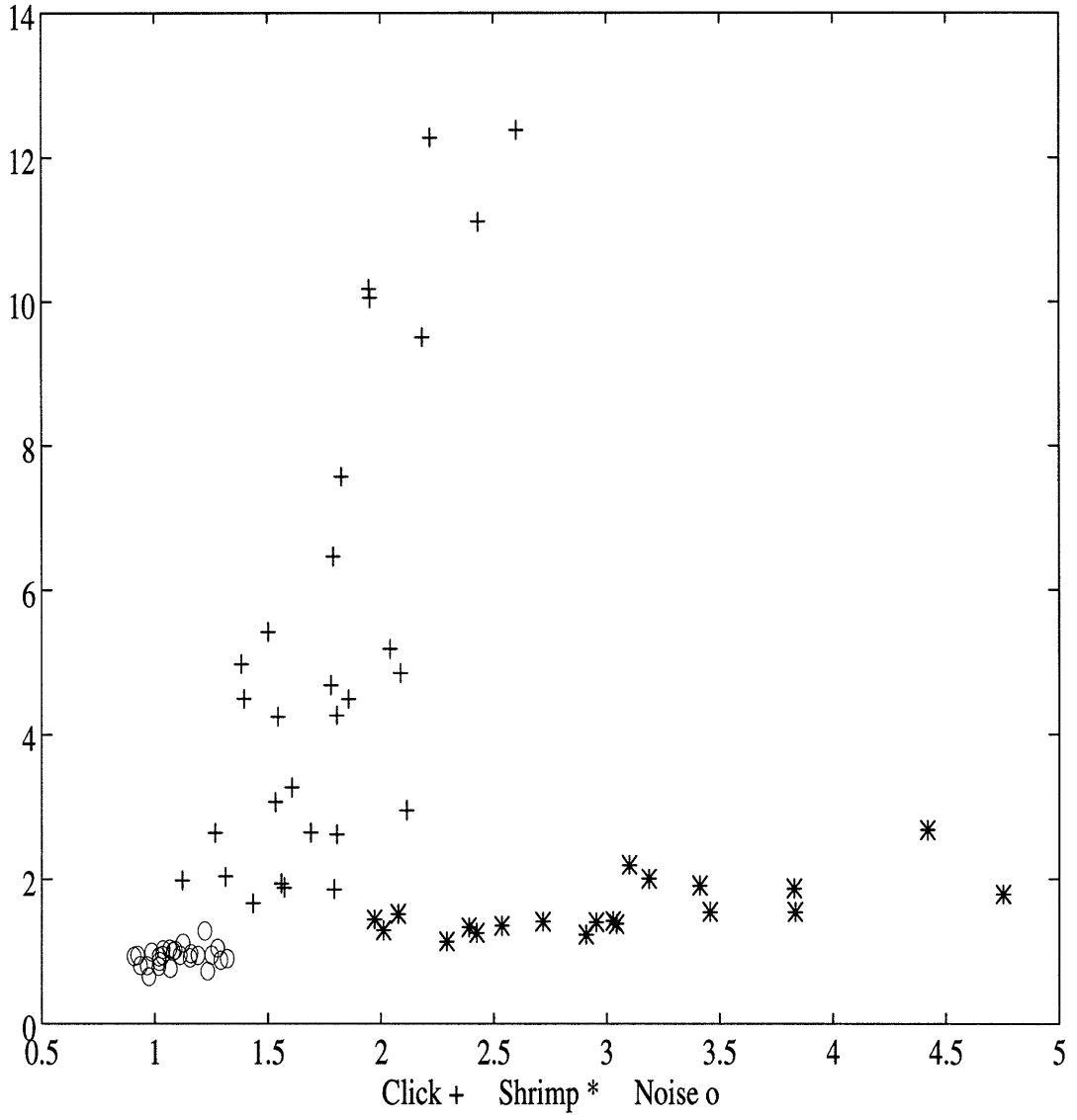


Figure 13: Noise normalized energies from bins b(4,1) and b(4,2) of the sample energy maps that make up the training set. b(4,2) vs b(4,1).

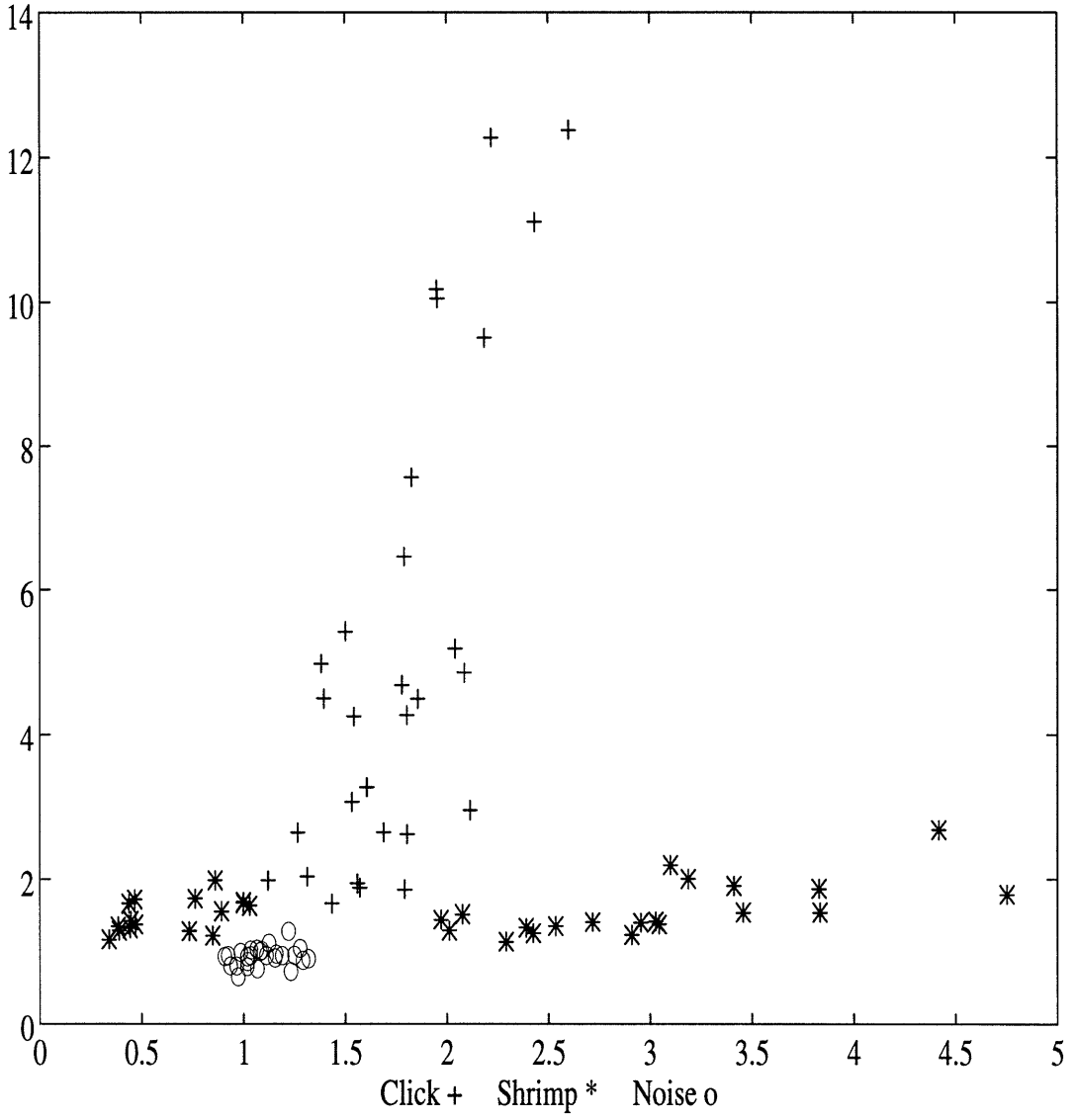


Figure 14: Noise normalized energies from bins b(4,1) and b(4,2) of the energy maps for the original data set of snapping shrimp, whale clicks, background noise and the new set of snapping shrimp.

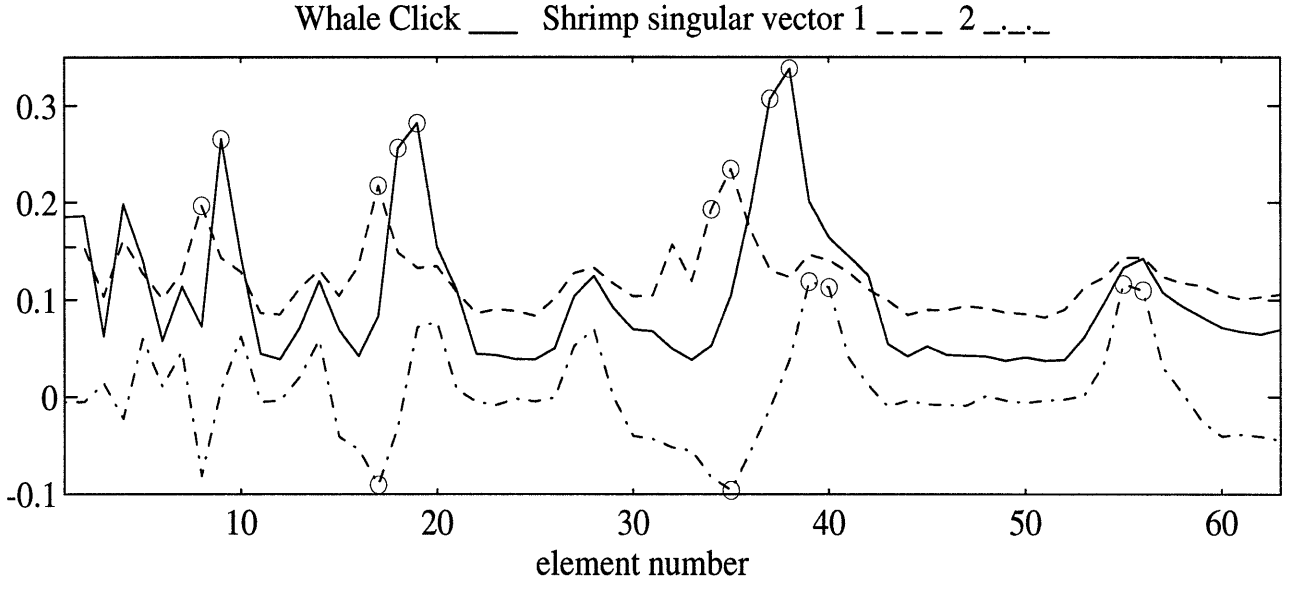
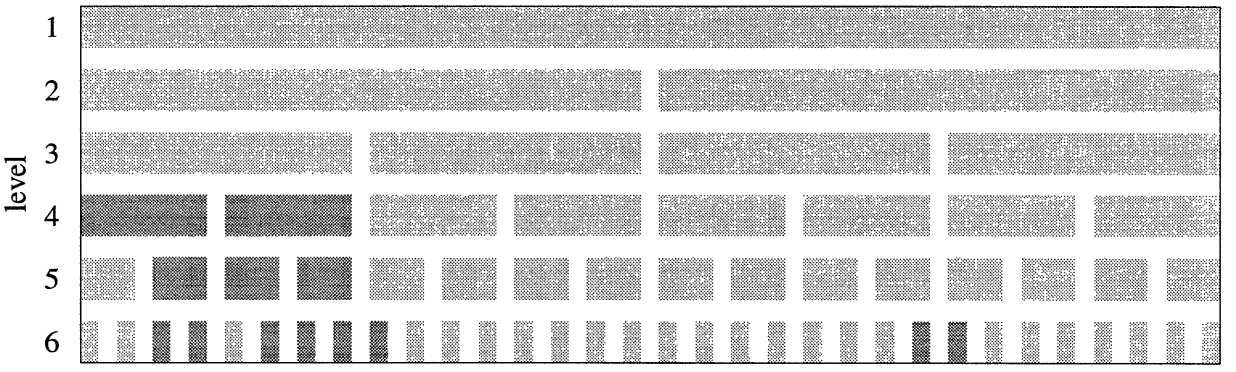
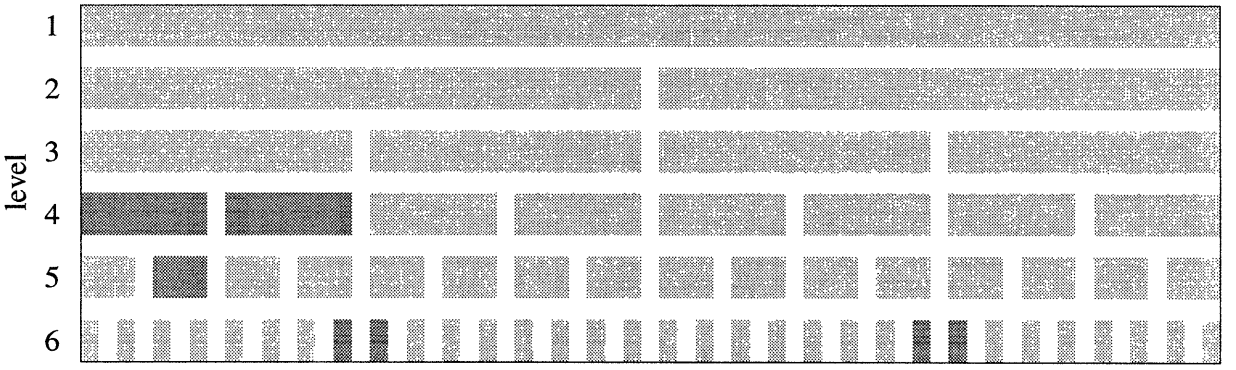


Figure 15: The 63 elements of the primary singular vector for whale clicks and the two primary singular vectors for snapping shrimp.



(a)



(b)

Figure 16: The significant bins are shaded. (a) The 13 bins found significant for whale clicks and all snapping shrimp data. (b) The 7 bins of 13 that do not exhibit parent-child redundancy.

References

[1] R. Coifman and M. Wickerhauser. "Entropy-based algorithms for best basis selection". *IEEE Trans. Info. Theory*, 38(2), March 1992.

[2] I. Daubechies. "Orthonormal bases of compactly supported wavelets". *Commun. Pure Appl. Math.*, 41, Nov. 1988.

[3] M. Desai and D. Shazeer. "Acoustic transient analysis using wavlet decomposition". In *Proceedings of the IEEE conference on neural networks for ocean engineering*, Wash. DC, Aug. 1991.

[4] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis.* John Wiley and Sons, 1973.

[5] B. Friedlander and B. Porat. "Performance analysis of transient detectors bsed on a class of linear data transforms". *IEEE Trans. Info. Theory*, 38(2), March 1992.

[6] M. Frisch and H. Messer. "The use of the wavelet transform in the detection of an unknown transient signal". *IEEE Trans. Info. Theory*, 38(2), March 1992.

[7] P. Goupillaiud, A. Grossman, and J. Morlet. "Cycle-octave and related transforms in seismic signal analysis". *Geoexploration*, 23, 1985.

[8] A. Grossman and J. Morlet. "Decompositions of Hardy functions into square integrable wavelets of constant shape". *SIAM J. Math.*, 15, 1984.

[9] R. Learned. "Wavelet Packet Based Transient Signal Classification". Master's thesis, Massachusetts Institute of Technology, 1992.

[10] S. Mallat. "A theory for multiresolution signal decomposition: The wavelet representation". *IEEE Trans. Pattern Anal. Machine Intell.*, 11, July 1989.

[11] Y. Meyer. "Ondelettes et functions splines". In *Sem. Equations aux Derivees Partielles.*, Ecole Polytechnique, Paris, France, Dec 1986.

[12] NeuralWare Inc., Penn. Center West, Pittsburgh, PA 15276. *Neural Works Professional II PLUS and Neural Works Explorer*, 1990.

[13] J. Nicolas, A. Lemer, and D. Legitimus. "Identification automatique de bruits impulsifs en acoustique sous-marine par reseaux multi-couches". In *International workshop on neural networks and their applications*, Nimes, France, Nov. 1989.

[14] G. Strang. *Linear algebra and its applications.* Harcourt Biaci Jovanovich, 4th edition, 1988.

[15] M. Wickerhauser. "Lectures on Wavelet Packet Algorithms". Technical report, Washington University, Department of Mathematics, 1992.